Jet Users' Quickstart Training

7 November 2012

Agenda

- Schedule
- Overview of Changes
- Accessing Jet
- Moving Data to/from Jet
- Using Modules
- Compiling and Running Codes
- Submitting a Job
- Monitoring a Job
- Resource Utilization Inquiries
- Summary

Schedule

- Nov 5th
  - Migrate njet and ujet
- Nov 8th
  - Put fe5 into production in the new OS
- Nov 26th
  - Migrate tjet
  - Migrate fe2,fe3,fe4
  - Logins will be automatically directed to the new front end systems with new OS
  - Old cron services will be stopped
    - Node fe1 will be left behind so users can access their crontabs and migrate them to the new image
- Dec 5th
  - Migrate fe1 to new OS
  - At this point old cron entries will be gone.

What Changed?

- OS Upgrade to Centos 6.3
- All major packages upgraded
- Batch system upgraded, now based on Moab/ Torque

- No modules are loaded by default
- Modules reorganized
- Allocations going forward will be based on core-hours
- mpiexec now used to launch jobs
- ImageMagick no longer on default path

Why The Changes?

- The new sJet system required an OS upgrade
- The existing OS was getting old and harder to maintain
- We needed to update several key packages for performance and bug fixes
- We need to keep the OS consistent across all systems for maintainability purposes

Documentation

- Basic NESCC Web Site is

  rdhpcs.noaa.gov/boulder

- On the left side is a link to FAQ

  https://jetdocs.rdhpcs.noaa.gov/wiki/index.php/Main_Page

- The jetdocs wiki has been update and matches the new OS. All information regarding the OS has been removed. If you have questions about the old OS prior to the transition of tJet, please email the help system.

jetdocs

navigation

- Main page
- Community portal
- Current events
- Recent changes
- Random page
- Help

search

Go Search

toolbox

- What links here
- Related changes
- Upload file
- Special pages
- Printable version
- Permanent link

Main Page

The Jet RDHPCS Documentation

Below is information regarding the use of Jet. The original version of this document lives at https://jetdocs.rdhpcs.noaa.gov. If you are not at that site, please go there for the latest, up to date information. If you have any questions or comments regarding the material, please email the Help System.

NOTICE -- THESE PAGES ARE UNDERGOING ACTIVE CHANGES TO UPDATE THE INFORMATION FOR THE NEW OPERATING SYSTEM. 11/5/2012

For immediate migration and testing, use: ***Getting Started on sJet****.*

General	Help
About High Performance Computing at GSD	Getting an Account
About Jet	Logging In
System Information	Transferring Files
Policies and Important Information	How Do I Submit a Help Ticket
Jet User Meeting Slides	Migrating Your Data To Zeus
Getting Information About Your Account	Real-Time Experiment Application Form
	Common Questions and Problems
Getting Things Done	**Software**
Using Modules	System Software
Compiling Applications	Using the new mpirun tools
Editing on Jet	Using OpenMP and Hybrid OpenMP/MPI on Jet
Running and monitoring Jobs	Using the HSMS
Managing File System Allocations	
Starting a parallel application	
Profiling Applications	
Other Tips and Tricks	
Debugging a Parallel Application	
Starting Recurring Processes with Cron	

Getting Help

Email your questions to
rdhpcs.jet.help@noaa.gov

Agenda

- Schedule
- Overview of Changes
- Accessing Jet
- Moving Data to/from Jet
- Using Modules
- Compiling and Running Codes
- Submitting a Job
- Monitoring a Job
- Resource Utilization Inquiries
- Summary

Accessing Jet

- Accessing Jet is the same as before
  - ssh either to jet.rdhpcs.noaa.gov or jetrsa.rdhpcs.noaa.gov
- Until Nov 26th
  - Login, then ssh to any host out of fe5, fe6, fe7, fe8
  - Hit Ctrl-C and select a host
- After Nov 26th
  - Do nothing, you will be redirected to a random-host (as is done today)

```
abbeyroad:~ ctierney$ ssh -X -L48765:localhost:48765 jet.rdhpcs.noaa.gov
********************************************************************************
*                                WARNING!                                      *
********************************************************************************
* This is a United States Government computer system, which may be      *
* accessed and used only for official Government business by authorized *
* personnel.  Unauthorized access or use of this computer system may    *
* subject violators to criminal, civil, and/or administrative action.   *
*                                                                       *
* All information on this computer system may be intercepted, recorded, *
* read, copied, and disclosed by and to authorized personnel for        *
* official purposes, including criminal investigations.  Access or use  *
* of this computer system by any person, whether authorized or          *
* unauthorized, constitutes consent to these terms.                     *
********************************************************************************
ctierney@jet.rdhpcs.noaa.gov's password:
Permission denied, please try again.
ctierney@jet.rdhpcs.noaa.gov's password:
Permission denied, please try again.
ctierney@jet.rdhpcs.noaa.gov's password:
Last login: Tue Nov  6 21:33:23 2012 from 137.75.201.5
You will now be connected to the default jet front-end system.
Hit ^C within 5 seconds to select another host.
Menu:
=====
Select a host.  Enter the hostname, or a unique portion of a hostname:
Hostname        IP Address
fe1     140.172.21.194
fe2     140.172.21.195
fe3     140.172.21.196
fe4     140.172.21.197
fe5     140.172.21.198
fe6     140.172.21.199
fe7     140.172.21.205
fe8     140.172.21.206

Enter hostname here: fe5

Attention user:

A port-tunnel has been established for SCP data transfers
on port 14814 to host jet.rdhpcs.noaa.gov.
```

Agenda

- Schedule
- Overview of Changes
- Accessing Jet
- Moving Data to/from Jet
- Using Modules
- Compiling and Running Codes
- Submitting a Job
- Monitoring a Job
- Resource Utilization Inquiries
- Summary

Transferring Data To/From Jet

- Transferring data has not changed. See:

https://jetdocs.rdhpcs.noaa.gov/wiki/index.php/Transferring_Files

Agenda

- Schedule
- Overview of Changes
- Accessing Jet
- Moving Data to/from Jet
- Using Modules
- Compiling and Running Codes
- Submitting a Job
- Monitoring a Job
- Resource Utilization Inquiries
- Summary

Using Modules

- Modules allow the PATH and environment variables to be manipulated so that multiple versions of software can be supported
- There are no longer any default modules loaded on the system. You must select them every time
  - You can put your selections in .profile or .cshrc if you wish

Modules - Documentation

- See:
  - https://jetdocs.rdhpcs.noaa.gov/wiki/index.php/Using_Modules
- For sh, ksh, or bash see:
  - https://jetdocs.rdhpcs.noaa.gov/wiki/index.php/Running_and_monitoring_Jobs
    - See: Additional requirements for batch scripts using ksh, sh, or bash

Module Dependencies

- Often modules are dependent on other modules
- To ensure that the dependencies are satisfied, modules are only available when their dependencies are satisfied
- Typically dependencies are due to compiler or MPI requirements

```
# module avail

---------------------------- /apps/Modules/versions -----------------------------------
3.2.9

------------------------ /apps/Modules/3.2.9/modulefiles --------------------------------
bbcp/12.01.30.01.0(default)     intel/12.1.4(default)          pgi/12.5-0(default)
cnvgrib/1.2.3(default)          intel/12.1.5                   rocoto/1.0.2
cuda/4.2.9(default)             lahey/8.10b(default)           rocoto/1.0.3(default)
dot                             module-cvs                     szip/2.1
gempak/6.7.0-gfortran(default)  module-info                    totalview/8.9.2-2(default)
grads/2.0.1(default)            modules                        udunits/1.12.11
hpss                            ncl/6.0.0                      udunits/2.1.24(default)
idl/8.2(default)                nco/4.1.0                      use.own
imagemagick/6.2.8(default)      ncview/2.1.1(default)          wgrib/1.8.1.0b(default)
intel/11.1.080                  null                           wgrib2/0.1.9.6a(default)


#
```

- Notice that important tools such as netcdf, hdf4, or mvapich are not available.

```
# module load intel
# module avail

---------------------------- /apps/Modules/versions 3.2.9 -----------------------------

------------------------ /apps/Modules/3.2.9/modulefiles --------------------------------
bbcp/12.01.30.01.0(default)    intel/12.1.4(default)          pgi/12.5-0(default)
cnvgrib/1.2.3(default)         intel/12.1.5                   rocoto/1.0.2
cuda/4.2.9(default)            lahey/8.10b(default)           rocoto/1.0.3(default)
dot                            module-cvs                     szip/2.1
gempak/6.7.0-gfortran(default) module-info                    totalview/8.9.2-2(default)
grads/2.0.1(default)           modules                        udunits/1.12.11
hpss                           ncl/6.0.0                      udunits/2.1.24(default)
idl/8.2(default)               nco/4.1.0                      use.own
imagemagick/6.2.8(default)     ncview/2.1.1(default)          wgrib/1.8.1.0b(default)
intel/11.1.080                 null                           wgrib2/0.1.9.6a(default)

------------------- /apps/Modules/default/modulefamilies/intel ---------------------------
hdf4/4.2.7              mvapich2/1.8(default)    netcdf4/4.2.1.1(default)
hdf5/1.8.9(default)     netcdf/3.6.3(default)    openmpi/1.6
mvapich2/1.6            netcdf4/4.2

#
```

- Now hdf4, hdf5, netcdf, and mvapich are available.
- What about parallel hdf5?

```
# module load mvapich2
# module avail

---------------------------- /apps/Modules/versions -----------------------------------
3.2.9

----------------------- /apps/Modules/3.2.9/modulefiles --------------------------------
bbcp/12.01.30.01.0(default)    intel/12.1.4(default)          pgi/12.5-0(default)
cnvgrib/1.2.3(default)         intel/12.1.5                   rocoto/1.0.2
cuda/4.2.9(default)            lahey/8.10b(default)           rocoto/1.0.3(default)
dot                            module-cvs                     szip/2.1
gempak/6.7.0-gfortran(default) module-info                    totalview/8.9.2-2(default)
grads/2.0.1(default)           modules                        udunits/1.12.11
hpss                           ncl/6.0.0                      udunits/2.1.24(default)
idl/8.2(default)               nco/4.1.0                      use.own
imagemagick/6.2.8(default)     ncview/2.1.1(default)          wgrib/1.8.1.0b(default)
intel/11.1.080                 null                           wgrib2/0.1.9.6a(default)

----------------- /apps/Modules/default/modulefamilies/intel ---------------------------
hdf4/4.2.7               mvapich2/1.8(default)    netcdf4/4.2.1.1(default)
hdf5/1.8.9(default)      netcdf/3.6.3(default)    openmpi/1.6
mvapich2/1.6             netcdf4/4.2

----------- /apps/Modules/default/modulefamilies/intel-mvapich2/1.8 --------------------
hdf5parallel/1.8.9(default)          netcdf4-hdf5parallel/4.2.1.1(default)
netcdf4-hdf5parallel/4.2

#
```

- Now the modules that have both compiler and MPI dependencies are available for use.

Agenda

- Schedule
- Overview of Changes
- Accessing Jet
- Moving Data to/from Jet
- Using Modules
- Compiling and Running Codes
- Submitting a Job
- Monitoring a Job
- Resource Utilization Inquiries
- Summary

Compiling Codes

- This works the same as the Old OS
- Select a compiler via modules
- Select a MPI stack via modules (if necessary)

Compiling Applications			
Compiler	**Fortran**	**C**	**C++**
Intel	ifort	icc	Icpc
Portland Group (PGI)	pgfortran	pgcc	pgCC
Lahey	lf95	gcc	gcc
Parallel Applications*	mpif90	mpicc	mpiCC

* Applies to all compilers and MPI stacks

Running an MPI Program

- Two MPI Implementations: Mvapich2 and OpenMPI
- mvapich2

```
mpiexec –np <numMPIprocs> <executable>
```

- OpenMPI

```
To Be Determined …
```

Agenda

- Schedule
- Overview of Changes
- Accessing Jet
- Moving Data to/from Jet
- Using Modules
- Compiling and Running Codes
- Submitting a Job
- Monitoring a Job
- Resource Utilization Inquiries
- Summary

Submitting a job

- qsub is used to submit jobs to the batch system
- qsub options

```
-A project (required)
-l procs=<number_of_cores>
-l walltime=<hh:mm:ss> or walltime=<seconds>
-l vmem=<memsize> (default is 1.2GB/process)
-q <queue_name> (default is batch)
-p partition=<Partition Name>
```

- Alternate method to specify resources

```
-l nodes=<nodes>:ppn=<proc_per_nodes>
```

Partitions on Jet

- To control job placement, the concept *partition* is used.

SGE Parallel Environment	Torque Partition
ncomp	ncomp
ncomp + faa reservation	nfaa
hnewcomp	hjet
Hfip	njet:tjet:ujet
thfip	tjet:ujet
nhfip	njet

Note: There is no serial queue! Just ask for 1 core (-lprocs=1) and the system will take care of the rest.

Jet Queues

Queue	Min Cores	Max Cores	Max Wallclock	Description
batch	1	2048	8:00:00	Default queue for jobs
service	1	1	24:00:00	Jobs will be run on front end nodes that have external network connectivity. Useful for data transfers, HSMS access, or access to external resources like databases. **You must also select the partition service to use the service queue**
debug	1	2048	00:30:00	Highest priority queue, used for quick turnaround when debugging applications.
novel	2048	10080	8:00:00	Queue for running novel or experimental jobs where nearly the full system is required. Permission may be needed to use this queue.
urgent	1	2048	8:00:00	Queue for rare jobs that need to start ASAP. Only one job per project can be running at a time in this queue.

Notes about the Jet Queues

- The queues have been implemented to match program requirements
- Current core maximums are for all projects
  - We will revisit this policy as usage ramps up
- System usage will be controlled through core-hour allocations per project

Useful Options and Variables

- To have a job start in your current directory

  -d .

  or

  cd $PBS_O_WORKDIR (In your batch script)

- Some Useful Job Environment Variables

  $PBS_JOBID - The jobid of the currently running job

  $PBS_O_WORKDIR - The directory from which the batch script was submitted

  $PBS_QUEUE - The assigned queue for this job

  $PBS_NP - The number of tasks assigned to this job

Examples

- Submit a job with a 4-hour limit, 32 cores, project nesccmgmt, and urgent queue
  - qsub –A nesccmgmt -l walltime=4:00:00,procs=32 –q urgent job.csh
- Submit a job with a 2-hour limit, 16 processes each with 6 threads (OpenMP/MPI hybrid), project fim, and batch (default) queue. Embed commands in script
  - `qsub fimjob.sh`
  - fimjob.sh:

```
#!/bin/sh –login
# Options use the PBS sentinel
#PBS –A fim
# Indicate 8 nodes with 2 MPI tasks/node (use 6 OMP threads/task)
#PBS –l nodes=8:ppn=2
#PBS –l walltime=7200
#PBS –l partition=njet:tjet:ujet
export OMP_NUM_THREADS=6
```

- See: https://nesccdocs.rdhpcs.noaa.gov/wiki/index.php/Running_and_Monitoring_Jobs

Submitting an Interactive Job

- To submit an interactive job use the –I and –X options

```
qsub –A fim -I -X –lpartition=tjet:ujet:njet –l procs=36,walltime=1:00:00
```

  - After the jobs works through the queue, an interactive prompt will appear
- Starting TotalView
  - At an interactive prompt enter the following (make sure you used –X on your ssh and msub):

```
# module load totalview
# totalview
```

- Starting TotalView via mpiexec

```
# module load totalview
# totalview mpiexec -np <numMPIprocs> <executable>
```

Agenda

- Schedule
- Overview of Changes
- Accessing Jet
- Moving Data to/from Jet
- Using Modules
- Compiling and Running Codes
- Submitting a Job
- Monitoring a Job
- Resource Utilization Inquiries
- Summary

Monitoring a job

- Show all jobs in the queue

```
showq
```

- Show all jobs belonging to only you

```
showq –u Christopher.W.Harrop
active jobs------------------------
JOBID              USERNAME      STATE PROCS   REMAINING            STARTTIME


0 active jobs           0 of 27708 processors in use by local jobs (0.00%)
                         0 of 2310 nodes active      (0.00%)

eligible jobs----------------------
JOBID              USERNAME      STATE PROCS     WCLIMIT            QUEUETIME

5646               Christop       Idle     1    00:05:00  Thu Jan 26 16:58:52

1 eligible job

blocked jobs-----------------------
JOBID              USERNAME      STATE PROCS     WCLIMIT            QUEUETIME


0 blocked jobs

Total job:  1
```

Monitoring a Job

- Using qstat will provide current status (showq can be up to 60 seconds old)
  - qstat –u John.Smith shows only jobs associated with user John Smith (note mixed case)
  - qstat (with no options) gives all jobs in all states
  - Many, many options (see man qstat)

Deleting a job

- `qdel <jobid>`
- `qdel –k <jobid>`
  - Please only use when qdel fails to terminate the job
  - Really kill it!

Agenda

- Schedule
- Overview of Changes
- Accessing Jet
- Moving Data to/from Jet
- Using Modules
- Compiling and Running Codes
- Submitting a Job
- Monitoring a Job
- Resource Utilization Inquiries
- Summary

Using Resources on Jet

- The old method of maximum cores per project to throttle usage is going away
- Moving to core-hour allocations
- Allocations will be on a monthly basis
  - Use it or lose it
- RDHPCS allocations specified by HPC program
- We are working with HFIP program to define allocations for projects

Using Allocations

- Using the allocations will be transparent
- When you exhaust your allocations, jobs can continue to run in windfall queue (low priority)
- Windfall queue can not be directly requested
  - Use your allocations first
- Your account will be assigned access to a specific system (sjet, njet, etc)

Resource Utilization Queries: account_params

```
# account_params
Account Params -- Information regarding project associations

User: ctierney
     Project: sepp
     Initial Allocation: N/A

          Directory: /lfs1/projects/sepp   DiskInUse=1702 GB, Quota=2000 GB

     Project: jetmgmt
     Initial Allocation: 1000.00
          Allocation: Id Name            Available Allocated PercentUsed
          Allocation: -- --------------- --------- --------- -----------
          Allocation: 1  Project=jetmgmt   1000.00   1000.00 0.00

          Directory: /lfs1/projects/jetmgmt   DiskInUse=13229 GB, Quota=15000 GB
          Directory: /lfs2/projects/jetmgmt   DiskInUse=271746 GB, Quota=300000 GB
          Directory: /pan2/projects/jetmgmt   DiskInUse=279968 GB, Quota=668616 GB

Allocation Terms:
                  Name -- Name of allocated project
                Amount -- Total Allocation
              Reserved -- Reserved requests from current jobs in queue.
               Balance -- Total allocation remaining for new jobs.
             Available -- Amount currently available for new jobs.
     Initial Allocation -- Current Month's Allocation.

           All amounts are in core-hours.

           Note: If the allocation of the disk space seem unusually large,
                 it is most likely because no allocation is defined for your project.
```

Agenda

- Schedule
- Overview of Changes
- Accessing Jet
- Moving Data to/from Jet
- Using Modules
- Compiling and Running Codes
- Submitting a Job
- Monitoring a Job
- Resource Utilization Inquiries
- Summary

Review Of Changes

- Use mpiexec to launch mpi applications
  - There is no real standard is to which mpiexec or mpirun should be supported. In the latest stacks, mpiexec does the right thing so we are going to use that.
- ImageMagick (convert, display) is no longer on the path
  - The ImageMagick package from CentOS is broken, and Redhat has not fixed it up stream.
  - ImageMagick now installed outside the OS and is accessible via **module load ImageMagick**

Review of Changes, cont

- Use **account_params** to find information regarding your projects

```
[John.Smythe@fe1 ~]$ account_params
Account Params -- Information regarding project associations

User: John.Smythe
     Project: sepp
          Allocation: Id  Name     Amount       Reserved Balance      CreditLimit Available
          Allocation: --- -------- ------------ -------- ------------ ----------- ------------
          Allocation: 106 sepp        100000000        0    100000000           0    100000000

          Directory: /lfs1/projects/sepp   DiskInUse=0 GB, Quota=5000 GB

     Project: nesccmgmt
          Allocation: Id Name      Amount       Reserved Balance      CreditLimit Available
          Allocation: -- --------- ------------ -------- ------------ ----------- ------------
          Allocation: 84 nesccmgmt     99935776        0     99935776           0     99935776

          Directory: /pan2/projects/nesccmgmt   DiskInUse=83768 GB, Quota=211500 GB
          Directory: /lfs2/projects/nesccmgmt   DiskInUse=137748 GB, Quota=211500 GB
```

Review of Changes, cont

- No modules are loaded by default
- Modules only are available for loading if their dependencies are satisfied

Getting Help

Email your questions to
rdhpcs.jet.help@noaa.gov

Thank You!

now…
Open Forum